

Secure Operations on Tree-Formed Verification Data

Andreas U. Schmidt and Andreas Leicher
Novalyst IT AG
Robert-Bosch-Straße 38, 61184 Karben, Germany
Email at: <http://andreas.schmidt.novalyst.de/>

Yogendra Shah and Inhyok Cha
InterDigital Communications, LLC
781 Third Avenue, King of Prussia, PA 19406
Email: {yogendra.shah,inhyok.cha}@InterDigital.com

Abstract—We define secure operations with tree-formed, protected verification data registers. Functionality is conceptually added to Trusted Platform Modules (TPMs) to handle Platform Configuration Registers (PCRs) which represent roots of hash trees protecting the integrity of tree-formed Stored Measurement Logs (SMLs). This enables verification and update of an inner node of an SML and even attestation to its value with the same security level as for ordinary PCRs. As an important application, it is shown how certification of SML subtrees enables attestation of platform properties.

I. INTRODUCTION

The process of building trust in computing platforms follows a unique, common pattern [1]. All components of the platform are measured by a protected entity on the platform before they are loaded and executed. The generation of a chain of trust, which extends without gaps from system boot up to the current system state is an important concept for a Trusted Computing System. To prevent unmonitored execution of code between measurement and actual execution, every component is required to measure and report the following component before executing it, while this measurement process is protected by the root of trust for measurement. *Verification data* is compiled from the measurement values by a protected operation and stored in protected storage. The verification data identifies, after completion of secure start up, the platform's state uniquely. Specified by the Trusted Computing Group (TCG), the most important embodiments of these processes are *authenticated boot* [2] and *secure boot* [3]. Secure boot includes a local verification and enforcement engine that lets components start only if their measurements are equal to trusted reference values.

In [4], a modification of the extend operation of TPMs was described which allows a verification data register, i.e., a PCR, to protect the root of a Merkle hash tree, which is stored in a *tree-formed SML*. This exposes a new TPM command, `TPM_Tree_Extend`, which assures that a sequence of measurements of system components and/or data are organised into a binary tree of which a designated verification data register is the root. Tree-formed SML extends the verification data from the root register to a complex data structure. It has various usages, for instance the efficient search for *failed* components, i.e., leaf measurements with undesired values.

In the present paper, we add more trusted functionalities to operate on tree-formed verification data. We show how inner nodes of a tree-formed SML with its root protected in

a verification data register, can be verified for integrity, and updated with a new value, in a controlled way maintaining the overall security level. Finally, we introduce a variant of the `TPM_Quote` command for inner tree nodes, which attests to their integrity precisely as `TPM_Quote` does for an ordinary PCR's value. With the defined set of commands, the integrity measurement functionality of a TPM is complemented by a comprehensive set of commands operating with tree-formed PCRs and SMLs. Using them, tree-formed verification and validation data can be used with far more flexibility and expressiveness than linearly chained TPM PCRs and SMLs.

Section II introduces a basic system model and notation. Section III defines the TPM command extensions described above, and also proposes some pertinent structural extensions and basic usage categories for such operations. As a central use case for the introduced commands, Section IV exhibits a protocol for certification of the root node of a subtree in a tree-formed SML by a trusted third party. We conclude with Section V with a brief assessment of the tree-formed platform validation approach and an outlook to future work. The reader is referred to [4] for an overview of related work. More related work is discussed as the concepts are developed.

II. PREPARATIONS

In this section, the minimal necessary elements and capabilities of a platform, which are required subsequently, are described. While we are leaning on TCG-nomenclature and some concepts, it will be clear from these minimal requirements, that the concepts developed in the following sections are not restricted to platforms and secure hardware elements adhering to TCG standards, e.g., TPMs [5] and systems designed according to the PC Client specification [2].

A. System Model

The tree-formation variant of the extend operation defined in [4], operates inside a TPM, takes only a single measurement value as input, and is otherwise inert with regard to the system outside the TPM. This is not the case for the update function which is introduced in the present paper. The latter operates on a certain number r of verification data registers V , $\mathcal{V} \stackrel{\text{def}}{=} \{V_1, \dots, V_r\}$ protected inside a TPM, and on the hash tree data stored outside the TPM in less protected storage. That is, the hash tree contained in the Stored Measurement Log (SML) is managed by a Trusted Software Stack (TSS) which is authorised to access the TPM functions necessary for the

update operations. TSS calls TPM via an authorised, integrity-protected command interface. Note that, while we use TCG parlance for practical reasons, the concepts presented here and in [4] are not restricted to a TCG TPM and platform. We only assume a hardware-protected set of verification data registers and an extend operation. The latter is defined by the ordinary TPM extend operation

$$V \leftarrow V \diamond m \stackrel{\text{def}}{=} H(V \| m), \quad (1)$$

where V denotes a verification data register, H is a collision-resistant hash function (SHA-1 in case of the TPM), and $m = H(\text{data})$ is a measurement value. In the following \diamond is used liberally with arbitrary registers V as arguments, where no confusion can arise.

B. Conventions

We assume that the SML contains a binary tree of depth d resulting from a binary one-way operation, such as the Merkle hash tree [6], [7] produced by the `TPM_Tree_Extend` command introduced in [4]. Natural coordinates for inner nodes and leaves are binary strings of length $1, \dots, d$, where the length ℓ of the string is the level in the tree on which the node resides. Let \mathbf{n} be an inner node or leaf and write $\mathbf{n} \sim \langle \mathbf{n} \rangle = (n_1, \dots, n_\ell) \in \{0, 1\}^{\times \ell}$ for the binary representation of the coordinates of \mathbf{n} . Let $\langle \mathbf{n} \rangle_k = n_k$, $k = 1, \dots, \ell$, be the k -th digit of $\langle \mathbf{n} \rangle$. Where no confusion can arise, we identify a node with its value (e.g., 160-Bit hash value) in the SML, while distinguishing it from its coordinate. Otherwise we write $\mathbf{n} = (n, \langle \mathbf{n} \rangle)$ for the value-coordinate pair of a node.

The **trace** \mathbf{T} of \mathbf{n} is the ordered list of all inner nodes on the path from \mathbf{n} to the root, including \mathbf{n} , i.e.,

$$\mathbf{T}(\mathbf{n}) = (\mathbf{t}_1, \dots, \mathbf{t}_\ell), \text{ where } \mathbf{t}_k \sim (n_1, \dots, n_k). \quad (2)$$

The natural partial order of nodes is written as $\mathbf{m} \leq \mathbf{n}$, which is equivalent to $\mathbf{n} \in \mathbf{T}(\mathbf{m})$. The partial order extends to sets M, N of nodes by setting $M \leq N$ iff $\forall \mathbf{m} \in M: \exists \mathbf{n} \in N: \mathbf{m} \leq \mathbf{n}$.

The **reduced tree** \mathbf{R} of \mathbf{n} is the list of all siblings of its trace. This is readily expressed in natural coordinates.

$$\mathbf{R}(\mathbf{n}) = (\mathbf{r}_1, \dots, \mathbf{r}_\ell), \text{ where } \mathbf{r}_k \sim (n_1, \dots, \neg n_k), \quad (3)$$

where \neg denotes binary negation.

We use the hash chain operation $x \diamond y \stackrel{\text{def}}{=} H(x \| y)$, with fixed-length input hash values x, y , in a variant which makes argument order dependent on a binary parameter. We set, for $c \in \{0, 1\}$,

$$\langle x | c | y \rangle = \begin{cases} x \diamond y & \text{if } c = 1; \\ y \diamond x & \text{if } c = 0. \end{cases}$$

This *chiral Merkle-Damgård operation* is a version of an extend operation which allows to distinguish between left and right siblings in a tree and calculate their parent node in the correct order. Neglecting implementation issues, we assume that the (extended) TPM is capable of performing the operation $\langle \cdot | \cdot | \cdot \rangle$ internally.

In many cases, the hash tree stored in the SML may be incomplete, i.e., contain empty leaves and inner nodes, denoted by *nil*. For the consistent treatment of *nil* nodes in a Merkle hash tree, it is useful to assume that *nil* is a two-sided unit for the operation \diamond , i.e.,

$$x \diamond \text{nil} = \text{nil} \diamond x = x, \text{ and } \text{nil} \diamond \text{nil} = \text{nil}. \quad (4)$$

This is a re-interpretation of the usual TPM extend operation and can also be used to model a direct write to a $V \in \mathcal{V}$, by first resetting V to *nil* and then performing $V \diamond x$ for some value x . For the implementation of this convention, we may assume that *nil* is represented as a flag of verification data registers and the inputs and output of $\langle \cdot | \cdot | \cdot \rangle$. For a V , the *nil* flag may be set by a particular reset command. When *nil* is encountered as the input of an extend operation to a V , then logic of the TSS, or a TPM modification, may prevent execution of the extend and write to the PCR directly.

III. SECURE OPERATIONS WITH TREE NODES

This main section presents the operational extensions of a standard TPM to operate securely with tree-formed SMLs. The protection goal is to achieve the same assurance level for inner nodes and leaves of such an SML, as for a conventional verification data register value, protected in a PCR. We first describe the update of a root by a new node value, and then show further structural and command extensions for use with tree-formed verification data.

The strategy for a secure update of an inner node or leaf of a SML tree is as follows. First, the current value of that node needs to be verified for authenticity. This is done by recalculating the root of the tree, protected in a register V , (which is kept fixed in the remainder of the paper to simplify presentation) using the data contained in the reduced hash tree associated with the node. This verification must be a protected operation inside the TPM, called `TPM_Reduced_Tree_Verify_Load`. It also loads the verified reduced tree data into a set of verification data registers for use with the subsequent update operation `TPM_Reduced_Tree_Update`. This function takes a new value for the node to be updated, and uses the reduced tree data to update the parent nodes up to the root V . Both commands may be used separately for various purposes, e.g. standalone node integrity verification. For convenience, they may also be combined into a single node and root update command.

A. Verified Load of a Reduced Tree

Suppose \mathbf{n} is the node of an SML tree of depth d at level $\ell \leq d$ with root protected in a verification data register $V \in \mathcal{V}$. The first step to update V with a new value for \mathbf{n} , is to verify that the reduced tree $\mathbf{R}(\mathbf{n})$ is untampered in the SML. To maintain the security level of V , this verification needs to be performed by a TPM-protected operation as well. For this, TSS calls `TPM_Reduced_Tree_Verify_Load` with arguments $(\langle \mathbf{n} \rangle, \mathbf{n}, \mathbf{R}(\mathbf{n}))$. Choose $\ell + 1$ available registers from \mathcal{V} and call them B_1, \dots, B_ℓ , and V^* . Algorithm 1 shows how an SML node is verified and its reduced tree is loaded into a set of verification data registers.

The chiral extend used centrally in this algorithm ensures correct order of the child nodes in the calculation of their parent element on the trace of \mathbf{n} . The TSS obtains the calculated trace $\mathbf{T}(\mathbf{n})$ and the verification status as return values. Algorithm 1 requires $\ell + 1$ additional verification data registers.

Algorithm 1 TPM_Reduced_Tree_Verify_Load

Require: $B_1, \dots, B_\ell, V^* \in \mathcal{V}, (\langle \mathbf{n} \rangle, \mathbf{n}, \mathbf{R}(\mathbf{n}))$

Ensure: $B_1 \leftarrow \mathbf{r}_1, \dots, B_\ell \leftarrow \mathbf{r}_\ell, V^* \leftarrow \mathbf{n}$

▷ Initialise buffer with reduced tree and node to verify.

```

1: for  $k = \ell, \dots, 1$  do
2:    $V^* \rightarrow \text{TSS}$ 
3:    $V^* \leftarrow \langle B_k | \langle \mathbf{n} \rangle_k | V^* \rangle$ 
4: end for
5: if  $V \equiv V^*$  then
6:   return  $V$ 
7: else
8:   return “verification error”
9: end if
```

A simple variant of Algorithm 1 can operate using only a single verification data register, by processing the reduced tree sequentially, without storing the reduced tree inside the TPM. This auxiliary command TPM_Reduced_Tree_Verify may be useful for a plain verification of the SML by the TSS or another party. This is shown in Algorithm 2. The serialisation of $\mathbf{R}(\mathbf{n})$ required by this variant may be done using an input buffer realised in a software layer below the TSS, e.g., a TPM device driver, or by corresponding TPM internal logic.

Algorithm 2 TPM_Reduced_Tree_Verify

Require: $B \in \mathcal{V}, (\langle \mathbf{n} \rangle, \mathbf{n}, \mathbf{R}(\mathbf{n}))$

Ensure: $B \leftarrow \mathbf{n}$ ▷ Initialise buffer with node to verify.

```

1: for  $k = \ell, \dots, 1$  do
2:    $B \rightarrow \text{TSS}$ 
3:    $B \leftarrow \langle \mathbf{r}_k | \langle \mathbf{n} \rangle_k | B \rangle$ 
4: end for
5: if  $V \equiv B$  then
6:   return  $V$ 
7: else
8:   return “verification error”
9: end if
```

Like the original tree formation algorithm of [4], Algorithms 1 and 2 use non-standard operations, in particular chiral extend. Since the output target of chiral extend is always a verification data register, the operation can be implemented by loading the other argument into another verification data register (if it is not already there, as in Algorithm 1), and preceding the TPM-internal operation \diamond with a register swap, depending on the middle argument of chiral extend. This ensures the same protection level for all arguments.

B. Full Node Verification

The verification performed by algorithms 1 and 2 has a limited meaning since it only assures the integrity of the input node value with respect to the input reduced tree. In case of an integrity breach of the SML tree, more detailed information is desirable. We can obtain at least the tree level at which an integrity breach occurs, by performing the validation strategy via downward tree-traversal described in [4].

The command TPM_Tree_Node_Verify shown in Algorithm 3 returns the level at which an incorrect reduced tree and/or trace element first broke the integrity chain from the root to \mathbf{n} . It obviously does *not* allow to determine which sibling broke the chain. Further diagnostics would only be possible when a reference tree is available, see [4].

Algorithm 3 TPM_Tree_Node_Verify

Require: $B, C, D \in \mathcal{V}, (\langle \mathbf{n} \rangle, \mathbf{R}(\mathbf{n}), \mathbf{T}(\mathbf{n}))$

Ensure: $C \leftarrow V$ ▷ Initialise comparison register with root.

```

1: for  $k = 1, \dots, \ell$  do
2:    $B, D \leftarrow \mathbf{t}_k$  ▷ Load trace child into buffers
3:    $B \leftarrow \langle \mathbf{r}_k | \langle \mathbf{n} \rangle_k | B \rangle$ 
4:   if  $C \equiv B$  then
5:     ▷ Compare result with current parent, and if OK,
6:      $C \leftarrow D$ 
7:     ▷ make the trace element just verified the new parent.
8:   else
9:     return “verification error at level ” ||  $k, B$ 
10:  end if
11: end for
12: return “OK”
```

C. Root Update

Assume that TPM_Reduced_Tree_Verify_Load has been performed for a node \mathbf{n} which shall now be updated with a new value \mathbf{n}' . The command

Algorithm 4 TPM_Reduced_Tree_Update

Require: $\langle \mathbf{n} \rangle, \mathbf{n}'$

Ensure: $B_1 = \mathbf{r}_1, \dots, B_\ell = \mathbf{r}_\ell$

```

1:  $V \leftarrow \mathbf{n}'$ 
2: for  $k = \ell, \dots, 1$  do
3:    $V \rightarrow \text{TSS}$ 
4:    $V \leftarrow \langle B_k | \langle \mathbf{n} \rangle_k | V \rangle$ 
5: end for
6: return  $V$ 
```

TPM_Reduced_Tree_Update is called with argument \mathbf{n}' and may exclusively operate on the result of a determined, preceding TPM_Reduced_Tree_Verify_Load, which also fixes the node coordinate $\langle \mathbf{n} \rangle$ and the register V to be updated. To achieve this binding in a command sequence, various methods can be employed. First, the TPM can store and manage states and additional data for tree operations as described in Section III-E. Furthermore the sequence

of commands `TPM_Reduced_Tree_Verify_Load` and `TPM_Reduced_Tree_Update` should be bound cryptographically, for instance by rolling nonces as implemented by TPM protected OIAP/OSAP authorised command sessions [5, p. 60ff]. Finally, the two commands may be joined to a single update command `TPM_Tree_Node_Verified_Update`, with arguments $(\langle n \rangle, n, n', R(n))$. The node update commands return the new trace of n' and the new value of V , with which the TSS then updates the SML.

D. Verification Data Register States

With the association of verification data registers to certain nodes or roots of hash trees, and the associated commands `TPM_Tree_Extend` (defined in [4]), `TPM_Reduced_Tree_Verify_Load`, `TPM_Reduced_Tree_Update`, these registers V acquire statefulness. States of particular importance may be

- **Active Root (AR)** signifying a root of an SML tree currently under construction by the `TPM_Tree_Extend` operation.
- **Complete Root (CR)** signifying the root of a tree which is the completed result of the measurement of a number of components, i.e., `TPM_Tree_Extend` operations. AR can transition to CR when the tree is full, i.e., contains 2^d leaf measurements, or triggered by the TSS if it is desired to close a tree at a certain stage. A V in CR state should be protected against further updates with `TPM_Tree_Extend`, but may be accessed by `TPM_Reduced_Tree_Update` or even the normal `TPM_Extend` operation depending on policies and authorisation.
- **Tree Build (TB)** signifying a register used to build an active tree in another, AR register by the `TPM_Tree_Extend` operation.
- **Reduced Tree Node (RT)** signifying the result of `TPM_Reduced_Tree_Verify_Load`, i.e., one of the registers B_k . An RT V must be protected until the corresponding `TPM_Reduced_Tree_Update`, or another, authorised command occurs.

When more than one tree is managed, V s' states need to be associated to their respective trees, e.g., using Unique Identifiers (UIDs). Furthermore node coordinates may need to be stored for each or some register(s). These data could be held in a Verification Data Allocation Table (VDAT) inside the TPM, and managed by a Tree Data Management Unit (TDMU).

E. Quoting a Tree Node

TPM protected verification of a node value enables a new core semantic for platform validation by attestation. In particular, we can define a variant of `TPM_Quote` that attests to a certain node value. In its most elementary form, such a command `TPM_Tree_Node_Quote` is called with the same arguments as `TPM_Quote` plus the arguments

of `TPM_Reduced_Tree_Verify`. It then executes Algorithm 2, but additionally keeps a copy of n in another PCR V' . Upon success it executes `TPM_Quote` on V' . The receiver of such a quote should be made aware that the signature obtained is over an SML tree's inner node. One possibility would be to change the fixed string contained in the signed blob of the `TPM_Quote` command, which normally is "QUOT" [5, Part 3, line 2794 on page 161], to, say, "TREQUOT".

Attestation to a node value with this command provides to the node the same security assertion as quoting a verification data register (PCR) value with `TPM_Quote`. However, it bears the additional semantics that the value corresponds to some inner node of an SML tree, i.e., it effectively attests to the state of a certain subtree of which n is the root. To explicitly convey this semantics to a validator, additional data may be included in the AIK (Attestation Identity Key) signed attestation package, e.g., a string "Tree Node". The meaning of such an attribute can be sensibly strengthened, if it is only assigned by `TPM_Tree_Node_Quote` if the root register is a controlled SML root register resulting from `TPM_Tree_Extend` commands, i.e., it is in the CR state, cf. Section III-D. This control should be part of the quote generation.

For the validation of an attestation message, the validator needs only the value n of the quoted node $n = (n, \langle n \rangle)$. More information transfer to the validator is in principle not necessary, therefore the above description of `TPM_Tree_Node_Quote` follows a principle of minimal revelation. A variant of the command may also sign the node coordinate $\langle n \rangle$, if the position of the node in the SML tree matters for validation. Extended validation data transferred to a validator could also include the reduced tree of n and root verification data register, where this makes sense.

As a straightforward alternative, it would be possible to task the validator with the verification of the reduced tree. This approach is used in [8] to attest to the integrity of Web pages delivered by a Web server, and to bind this to an attestation of the server's state using the ordinary `TPM_Quote` command. This brings us to a variant realisation of `TPM_Tree_Node_Quote`, simply as follows. The command receives as arguments the node value n , the node values of $R(n)$, and a selector for the root V . The TPM signs¹ this (concatenated) data after controlling the CR state of V and with a "REDTREEQUOT" fixed string attribute.

The first and second realisation variant for quoting an inner node represent opposite possibilities, in the sense that the first puts verification load with the platform, while the second puts it with the validator. Therefore, both may have different domains of practical efficiency. For instance, many, distributed validating platforms as in M2M communication for the first, and many, distributed validators (such as the Web clients in [8]) for the second variant. But note that the second

¹Note that in [8], a workaround method is used to bind a reduced tree to a quote from a TPM, by inserting a hash of these additional data into the *nonce* input of the `TPM_Quote` command, which is normally used to guarantee freshness of the quote.

variant has, by principle drawbacks with regard to information revelation of the platform, since the validator is shown the complete state represented by V . This may be detrimental to privacy.

F. Baseline Applications

The various extensions of the TPM integrity measurement functionalities introduced in this paper and [4] can be grouped into the following categories. `TPM_Tree_Extend` is used in the (continuous) measurement process that builds a particular SML tree with PCR-protected root. `TPM_Reduced_Tree_Verify_Load`, `TPM_Reduced_Tree_Verify`, and ultimately `TPM_Tree_Node_Verify` are commands for platform-internal diagnostics. Apart from the usage of `TPM_Reduced_Tree_Verify_Load` as a preparation step to `TPM_Reduced_Tree_Update`, they may be used to verify certain properties of a platform, represented by SML subtrees, before other events can happen. `TPM_Reduced_Tree_Update` and `TPM_Tree_Node_Verified_Update` are used for the controlled update of subtrees. Particular usages of an inner node update operation are

- Update of a single system component. In this case the new value updates a leaf.
- Update of a system module represented by a subtree. In this case the root of the new subtree updates an inner node of the original tree.

Finally, `TPM_Tree_Node_Quote` is the command which makes tree-formed SML usable for validation of a platform by a remote party. It exhibits the key new element of validation using tree-formed data, namely the possibility to attest to subtrees representing only a defined part of the system state. A particular use case is described in the next section.

IV. SUBTREE CERTIFICATION

We come to a primary use case of the command extensions introduced in Section III. One of the biggest open problems of Trusted Computing is the association of semantics to platform attestation. Existing TCG specifications define a bilateral remote attestation in which executed code is measured when it gets loaded. The measurements are stored in PCRs as verification data, and the TPM attests to these data by signing them with a TPM protected Attestation Identity Key (AIK). Since a digest of a complete configuration is transmitted, the verifier needs to know all configurations of all machines (at all times, if system dynamics are considered). The transmitted data for validation thus lacks expressiveness to enable versatile and efficient remote platform validation. The need for semantic attestation was recognised early on in [9] and later in [10], who propose to restrict the scope of a single attestation to a virtualised subsystem with limited complexity, allowing for attestation of complex, dynamic, and high-level program properties. In [11] and [12], [13] “property,” respectively, “property-based attestation” (PBA) is proposed. PBA allows to assure the verifier of security properties of the verified

platform via a trusted third party (TTP), called *Subtree CA* (SCA). The SCA issues a certificate which maps the platforms configuration to the properties (in particular desired/undesired functionality) which can be fulfilled in this configuration. Essentially, PBA moves the infrastructural problem of platform validation to a SCA, similarly to, but extending the role of, the TCG’s privacy CA. Certification of subtrees is one way to fill the mentioned ideas with life. A related idea is that of hardware-supported updates [14], performed by a new, proposed TPM command, which re-seals data for another platform configuration based on an update certificate. Also, in [15] certificates over TPM-protected roots of hash tree are generated by a specific TPM command, to certify the state of certain protected counter objects. Differently from that, the present paper proposes a specific client-server protocol to obtain such a certificate from a trusted third party.

Let us fix some notions. In distinction to verification data, we call *validation data* all data that can be submitted to another party, the *validator*, and used to assess the trustworthiness of the state of the platform. The process of submission of validation data to the validator, for instance realised as remote attestation according to TCG, and evaluation thereof by the validator, is properly called *validation*. Validation data may often comprise verification data such as quoted verification data register (e.g., PCR) values. Validation may, beyond cryptographic verification of verification data, include policy evaluation and triggering of actions by the validator. See [1] for further discussion.

Tree-formed verification data and validation data associated to an SML tree, provide structured data which can be used complementary to the approaches above, to enhance the semantics of platform attestation. Here we present one fundamental method using tree-formed verification data to realise concepts related to PBA. Namely, it is shown how a SCA can replace a node in an SML tree with a meaningful, trusted statement — a *subtree certificate* — about the subtree of measured components of which latter node is the root. This process realises a *partial* validation of the platform by the SCA and results in a trusted assertion that is ingested in the available validation data of the platform. This can later be used toward another validator to validate the platform more fully. In the next subsection we describe a generic base protocol for subtree certification which is common for all conceivable variant realisations and use cases. After that, particularities and variants are considered in Subsections IV-B– IV-D.

A. Subtree Certification Protocol

Subtree certification is a process by which a Trusted Platform (the combination of TPM and TSS in our simplified model) obtains a certificate for the value of an inner node of a tree-formed SML from a SCA. For this, the platform submits a signed statement to the SCA, signifying that the node value is contained in an SML tree of with root value protected in a TPM verification data register. Based on this evidence, the SCA can issue a certificate with additional attributes about this node value to the platform, which is then ingested into

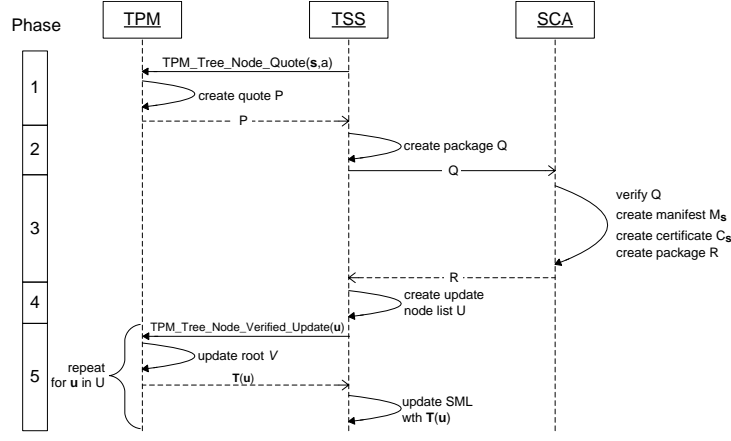


Fig. 1. Certification protocol for a subtree with root s .

the SML tree. The process essentially uses the protected tree operations introduced in Section III above.

We assume that the platform possesses an active AIK a , with certificate C_a issued by a trusted Privacy CA (PCA). We further assume that communication between the platform and the SCA is encrypted and integrity-protected to mitigate man-in-the-middle attacks. An inner node s is selected for certification. How this is done, shall concern us later in Subsection IV-D. In the protocol, we do not mention failure conditions and negative responses. With these preparations, the subtree certification protocol proceeds in five phases, as depicted in Figure 1.

Phase 1 creates a quote over s . For this, TSS calls `TPM_Tree_Node_Quote` with arguments $(\langle s \rangle, s, R(s), a)$ (note that Figure 1 shows only essential arguments for brevity) and receives back

$$P = \text{Sig}_a(s).$$

If the root of the tree, i.e. the register V is selected for certification, then `TPM_Quote` is to be used on V instead.

In phase 2, the TSS creates an attestation package Q . It contains all necessary information for the verifying SCA, at least

$$Q \subseteq \{P, s, C_a, a_{\text{pub}}\}.$$

(When the public part a_{pub} of a is not evident from C_a . Also, the value of s may be known to SCA and then be omitted from Q). More information may be included as necessary, for instance the node coordinate $\langle s \rangle$, when it is part of the quote. Q is sent (encrypted with a public encryption key of SCA and integrity-protected) to SCA. This phase is similar as in remote attestation specified by the TCG.

Phase 3 comprises the activities of SCA. First, SCA verifies Q by verifying the signature of P and tracing the certificate chain, up to the root certificate of the PCA, if necessary. If the SCA recognises s as a node value which it can certify, it creates a manifest M_s for it. This manifest may contain additional information about the platform state associated with the presence of the subtree with root s in the SML, such as a time stamp, a functionality of the platform, the identification of

a module combined from the loaded components represented by the leaf measurements of the subtree, or another platform property. The manifest is the validation data added by subtree certification which provides semantic meaning to the node value s to a validator. Now, SCA can create a certificate for s . This certificate, C_s , binds the properties represented by M to the platform, by binding it to the AIK a . This can be done essentially in two ways, namely

$$C_s = \begin{cases} \text{Sig}_{\text{SCA}}(M_s \| P) & \text{if } s \text{ is revealed;} \\ \text{Sig}_{\text{SCA}}(M_s \| \text{bind}(a)) & \text{if } s \text{ is concealed.} \end{cases} \quad (5)$$

In the first case, SCA signs the manifest and the AIK-signed node value, thus establishing an indirect binding to a . The binding of C_s to a can then be verified if the platform reveals the node value s . In the second option, the binding is achieved directly, by letting the SCA sign some data $\text{bind}(a)$ which uniquely identifies a , such as a 's public part, C_a , the serial number, or the fingerprint of C_a . In the semantics of Public Key Infrastructures, C_s is, by the binding, an attribute certificate associated with C_a . Finally, SCA creates a package R containing at least M_s and C_s , and $\text{bind}(a)$ in the second case, and returns it to the platform.

Phase 4 prepares the update of the SML with certain data derived from R . The SML update is an essential step to produce a binding association between the subtree certificate and the certified node's position $\langle s \rangle$ in the tree. Only this allows the platform to assert to a validator that the property attested by C_s and M_s is present in the platform's configuration. Various ways of SML update to bind C_s to the represented subtree are conceivable, each suited differently for particular use cases. This is discussed in Subsection IV-B, while we now state generic features of the SML update process.

A set $U = \{u_1, \dots, u_k\}$ of new node nodes (values and positions in the SML tree) is created with the following properties. First, it must hold $U \leq s$, so that only the subtree below s is touched by the update. This is necessary, since all old SML tree nodes $n \leq U$ strictly below U , i.e., $n \notin U$ are invalidated by the update, and can not be verified anymore with

respect to the updated root verification data register. Second, U is *dependency-free*, i.e.,

$$\nexists \mathbf{u}, \mathbf{u}' \in U: \mathbf{u} \leq \mathbf{u}'.$$

Dependency-freeness is the essential property ensuring consistency of the tree update by U with the one-way (upward) information flow embodied in Merkle hash trees. In particular it makes the update result independent of the order in which elements of U are processed.

Phase 5 is the SML tree update proper. Iterating over $\mathbf{u} \in U$, `TPM_Tree_Node_Verified_Update` is called with arguments $(\langle \mathbf{u} \rangle, \mathbf{n}, \mathbf{u}, \mathbf{R}(\mathbf{n}))$, where \mathbf{n} is the old SML node value at position $\langle \mathbf{u} \rangle$. This returns the new trace $\mathbf{T}(\mathbf{u})$ with which the TSS updates the SML. Executing the tree update in the way described above maintains a consistent security level for the SML and root verification data register. Namely, the operation \diamond is always executed inside the TPM. When U contains many elements, it may not be efficient to perform the update in the way described for Phase 5, since `TPM_Tree_Node_Verified_Update` would in such a case verify many overlapping reduced trees and thus incur redundancy in (complex) hash calculations. A more efficient update algorithm is described in Appendix A.

A variant of the subtree certification protocol could combine the roles of PCA and SCA for AIK, respectively, subtree certification in a single protocol run. An advantage would be that no explicit generation and verification of an AIK certificate C_a is necessary, because generation, activation, and use of the AIK are bound into one session. This combination of protocols is straightforward and left as an exercise to the reader.

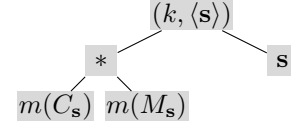
B. Certificate–Subtree Binding

Binding the received subtree certificate to the platform state means binding it to the tree-formed SML in the correct configuration, i.e., the position of the certified subtree’s root. As mentioned above, this is essential for meaningful subtree certificate-based validation in the context of an overall platform configuration. One particular goal of binding C_s to the SML tree is integrity protection, since, for instance, later replacement with a different certificate must be prevented. The binding can be achieved by updating parts of the tree with data which uniquely and verifiably identifies the subtree certificate. A wide range of data items can be produced from the subtree certificate and entered into the SML tree in various positions. Here, we describe some of the more sensible options.

In the simplest case the SML update may be trivial and U may be empty. This is only possible if C_s is composed by the first option of (5), revealing s . Then s can just be retained in the SML tree and whether the subtree below it is also retained depends on the use case. The binding association is via the actual node value s signed by C_s .

As another example, consider the case that all meaningful data concerning the platform property attested by the certificate should be protected by the updated tree, e.g., for forensic use. That is, the three data items s , M_s , and C_s shall enter the

update set. While the node value s is already in the correct data format, the other two are first processed to $m(M_s)$ and $m(C_s)$. The operation m can be the generation of a hash value by the platform’s Root of Trust for Measurement (RTM), or another appropriate one-way operation. If some data item already contains suitable, uniquely identifying data of the appropriate node value format, then it can be directly extracted and used as node update value. A particular example could be a certificate fingerprint contained in C_s . The three update nodes can then be configured in an update set to produce, for instance, the following configuration of updated nodes in the SML tree.



The root of the updated subtree is inserted in the old position of s and has the value $k = (m(C_s) \diamond m(M_s)) \diamond s$. This configuration provides independent integrity protection to the subtree certificate and manifest, and retains the old node value independently. In particular, attestation to the platform property represented by C_s can, in this configuration, still be done without revealing information about s , by quoting only the left inner node $*$ of the subtree.

Variants of certificate to subtree binding abound. The platform may also want to include (integrity protection values of) own generated data therein, for instance an internal time stamp from a secure clock. What makes sense depends ultimately on the use case.

C. Subtree Validation

For the attestation of the property represented by a subtree certificate to a validator, the platform can quote, using `TPM_Tree_Node_Quote`, any node in or above the updated subtree which protects the intended validation data, which comprises at least the manifest and the certificate proper. The platform will then submit validation data as necessary to the validator, at least all data needed for verification of the asserted property, again comprising at least M_s and C_s . Note that the validation data which is already protected by the submitted quote does in principle not require additional integrity protection in this.

One important point for the validator is to verify platform binding of the validation data. It is known that proving this property, i.e., that the validating platform is the same that performed subtree certification toward the SCA, is non-trivial [16]. The simplest way to achieve it is to use the *same* AIK, a , in subtree validation as in certification. The platform would then also submit a_{pub} , and if necessary also C_a as part of the validation data. Whether C_a is needed depends on the semantics of the subtree certificate, i.e., SC may already have checked the AIK certificate and C_s may state its veracity. According information can be placed in the manifest. Re-using the same AIK partially compromises privacy, and other methods to solve the problem may be worth further study.

D. Subtree Discovery

An important step for the practical use of subtree certification is the discovery of subtrees for which a platform can obtain certificates from a particular SCA. Without going into details of the interactions between platform, SCA, and validator, two categories of subtree discovery methods are described below. The first one places the workload of subtree discovery with the platform, while the second one assumes a “dumb” platform and places more load on the SCA.

1) *Active Discovery*: In this model, the SCA sends some *subtree discovery data* to the platform, in the simplest case a list of node values which it is ready to certify. The platform can search for these values in its SML tree and perform subtree certification for each identified node. This baseline procedure suggests various refinements, in particular enriching the discovery data and extending the discovery procedure to a negotiation protocol between platform and SCA. For instance, discovery data may contain root node positions as conditions on certifiable roots, which would, in the case of an SML produced in an authenticated boot process, correspond to the fact that the components loaded during the build of the latter subtree are loaded at a defined stage of the platform start up. Such conditions on absolute positioning of nodes may be difficult in practise for complex platforms whose configurations may change dynamically. More refined conditions could therefore also express relative positions of some, e.g., pairs of certifiable roots. The SC could state expressions saying “s is certifiable, if it is preceded by r” (i.e., r lies to the left of s in the ordered SML tree). This can be interpreted to the end that a certain functionality is operational on the platform only if another functionality was made operational before it.

A more fundamentally different variant of Model I is that the discovery data does not consist of subtree roots, i.e., inner nodes, but rather of leaf, i.e., measurement, values. A “bottom up” discovery procedure would require that the platform makes an “educated guess” about which inner nodes are certifiable, based on the received leaf measurement values which the SCA asserts to know as trusted values. A simplistic method is to find the set of span roots of subtrees whose leaves are all in the discovery data. The platform may then quote a subtree root and send it together with its SML subtree to the SCA. In general, the SCA will have to verify the SML subtree and decide if it is ready to certify that root, since this may still depend on the order of the leaves. In many cases, the platform may want to obtain a certificate for a subtree for which the particular SCA knows only *some* leaf values, i.e., the leaf set of the corresponding subtree has gaps with respect to the discovery data. If the platform has other trusted data, for instance RIM certificates [3] obtained from a party which the SCA trusts, the platform could submit these data in Phase 2 of the subtree certification, to aid SCA with its decision to certify the subtree root.

2) *Passive Discovery*: In the case that the device is not capable to perform a local discovery of subtrees, a model can be used which moves the computations to the SCA.

The platform selects an inner node n , with the aim to retrieve certificates from the SCA for any suitable subtrees below n . The node n could be equal to the root of the complete tree (V), in the case that the platform wants to get all certifiable nodes certified. The next two steps are the same as described in Phases 1 and 2 in section IV-A, i.e., the platform performs a `TPM_Tree_Node_Quote`, or `TPM_Quote`, respectively, if V was selected. Wrapped in an attestation package together with the SML subtree below the quoted root. The SCA receives this information and can then, using tree traversal techniques described in [4], verify the integrity of the tree and concurrently find one or multiple (disjoint) subtrees S_i with certifiable set of roots S . The SCA then iterates phase 3 of the protocol from IV-A, creating certificates for all $s_i \in S$. Since the protocol allows for the update of multiple nodes, incorporated into the update node list U , the update of all found subtrees can be done in a single protocol run. A variant could be for the platform to not send a `TPM_Tree_Node_Quote` or `TPM_Quote` in the first step, but only provide the SCA with the SML, starting from the selected node n . The SCA then searches for potential candidate subtrees to be certified and then requests the platform to provide a `TPM_Tree_Node_Quote` for the root nodes of the identified subtrees. This is a trade-off in the sense that it allows the platform to send the SML without having to perform cryptographic operations in advance. Nevertheless, the platform must provide appropriate quotes prior to the certification by the SCA to provide integrity protection for the sent SML.

V. CONCLUSION

Validation using tree-formed SMLs and verification data registers adds semantics to remote attestation. The possibility to attest to subtrees of an SML enables expressiveness far beyond conventional remote attestation. Tree-formed verification data is a promising way to substantiate other proposals to add semantics to platform attestation, e.g., association and of properties to validation data, as in PBA.

Further work shall pursue this promising direction and consider concrete architectures for platform validation with tree-formed verification and validation data — what we call *tree-formed validation* (TFV). One conceivable option is to efficiently organise a database of reference trees by an SCA and/or a validator in a way that allows for the modular building using subtrees of known component sub-structures, e.g., dependent programs loaded in sequence, or components with uniform associated security policies. Architectures and methods for subtree discovery, expression of dependencies between validated platform components, and management (updates, remediation) of platforms according to results of TFV are subjects of ongoing research and shall be discussed elsewhere.

APPENDIX A

EFFICIENT, SECURE NODE SET UPDATE

As mentioned in Subsection IV-A, the efficiency of updating a large set U of inner nodes using `TPM_Tree_Node_Verified_Update` depends on the overlap of the reduced trees of the elements of U , since many redundant hash operations for verification can occur. A bulk update strategy can be applied to improve the naïve algorithm of Phase 5 of the subtree certification protocol, using only the `TPM_Tree_Extend` command described in [4]. It rests on the observation that subsets of the update set U span subtrees which are independent of the old SML values, i.e., their roots depend only on nodes in U . Thus the roots of the trees spanned by such sets can be pre-calculated without expensive verification.

We first need some definitions. Assume $U = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a dependency-free update set. A node in the SML tree is called *U-intrinsic*, if a) it is an element of U , b) its sibling is in U , or c) its sibling is *U-intrinsic*. This recursive definition captures all nodes whose updated values depend only on U and not on SML nodes in the complement of U . The *span root* of a subset $V \subseteq U$ is the unique intersection point of the traces of all elements of V . The subtree *spanned* by a subset $V \subseteq U$ is the union of all traces of elements of V with all nodes strictly above the span root omitted. Now, the subset V is called *U-intrinsic* iff all elements of its spanned subtree are *U-intrinsic*.

With these settings, more efficient update of the SML with U is done as follows.

- 1) Identify the (mutually disjoint) *U-intrinsic* subsets $V_1, \dots, V_k \subseteq U$.
- 2) Iterate over $V_i, i = 1, \dots, k$.
 - a) Normalise the coordinates of elements of V_i by
 - i) truncating the prefix given by the coordinate of the span root of V_i , and
 - ii) post-fixing zeroes until all coordinates have equal length, the depth of the subtree spanned by V_i .
 - b) Order the elements of V_i alphabetically according to their normalised coordinates, producing an ordered list \tilde{V}_i .
 - c) Fill up all gaps (in the normalised coordinates) in \tilde{V}_i with *nil* values.
 - d) Select a free verification data register V' .
 - e) Sequentially use `TPM_Tree_Extend` on the elements of \tilde{V}_i with target V' .
 - f) Remove V_i from U .
 - g) Insert $(V', \langle \mathbf{v}_i \rangle)$ into U , where \mathbf{v}_i is V_i 's span root.
- 3) For the remaining elements of U , apply the normal update procedure of Phase 5 in Subsection IV-A using `TPM_Tree_Node_Verified_Update`.

ACKNOWLEDGEMENT

This work was funded by InterDigital, Inc. Special thanks go to Lawrence Case, Bob DiFazio, David Greiner, Louis

Guccione, Dolores Howry, Michael V. Meyerstein, and Sudhir Pattar, for many useful discussions and comments.

REFERENCES

- [1] A. U. Schmidt, I. Cha, and A. Leicher, "Scaling concepts between trust and enforcement," in *Trust Modeling and Management in Digital Environments: From Social Concept to System Development*, Z. Yan, Ed. IGI Global Publishing, 2010, pp. 20–57.
- [2] Trusted Computing Group, "TCG PC Client Specific Implementation Specification For Conventional BIOS," Version 1.20 FINAL Revision 1.00, July 2005, for TPM Family 1.2; Level 2.
- [3] —, "TCG Mobile Reference Architecture," Specification version 1.0 Revision 5, TCG, June 2008. [Online]. Available: <https://www.trustedcomputinggroup.org/>
- [4] A. Schmidt and A. Leicher, "Tree-formed Verification Data for Trusted Platforms," preprint, InterDigital, Inc., Novalyst IT, November 2009. [Online]. Available: <http://arxiv.org/abs/1007.0642>
- [5] —, "TPM Main," Specification Version 1.2 Level 2 Revision 103, TCG, July 2007. [Online]. Available: <https://www.trustedcomputinggroup.org>
- [6] R. C. Merkle, "A certified digital signature," in *Advances in Cryptology (CRYPTO '89)*, ser. LNCS, G. Brassard, Ed., no. 435. Springer-Verlag, 1989, pp. 218–238, republication of the 1979 original.
- [7] —, "Protocols for public key cryptosystems," in *Proceedings of the IEEE Symposium on Security and Privacy*, vol. 0. Los Alamitos, CA, USA: IEEE Computer Society, 1980, p. 122.
- [8] T. Moyer, K. R. B. Butler, J. Schiffman, P. D. McDaniel, and T. Jaeger, "Scalable web content attestation," in *Proceedings of the Twenty-Fifth Annual Computer Security Applications Conference, ACSAC 2009, Honolulu, Hawaii, 7-11 December 2009*. IEEE Computer Society, 2009, pp. 95–104.
- [9] V. Haldar, D. Chandra, and M. Franz, "Semantic remote attestation: a virtual machine directed approach to trusted computing," in *VM'04: Proceedings of the 3rd conference on Virtual Machine Research and Technology Symposium*. Berkeley, CA, USA: USENIX Association, 2004, pp. 29–41. [Online]. Available: <http://www.usenix.org/events/vm04/tech/haldar.html>
- [10] F. Baiardi, D. Cileia, D. Sgandurra, and F. Ceccarelli, "Measuring Semantic Integrity for Remote Attestation," in *Proceedings of the 2nd International Conference on Trusted Computing*. Springer, 2009, pp. 81–100.
- [11] J. Poritz, M. Schunter, E. V. Herreweghen, and M. Waidner, "Property Attestation—Scalable and Privacy-friendly Security Assessment of Peer Computers," IBM Zurich Research Laboratory, Rüschlikon, Switzerland, Research Report RZ 3548, 10 2004.
- [12] A. Sadeghi and C. Stübke, "Property-based attestation for computing platforms: caring about properties, not mechanisms," in *Proceedings of the 2004 workshop on New security paradigms*. ACM New York, NY, USA, 2004, pp. 67–77.
- [13] L. Chen, R. Landfermann, H. Löhr, M. Rohe, A. Sadeghi, and C. Stübke, "A protocol for property-based attestation," in *Proceedings of the first ACM workshop on Scalable trusted computing*. ACM New York, NY, USA, 2006, pp. 7–16.
- [14] U. Kühn, K. Kursawe, S. Lucks, A.-R. Sadeghi, and C. Stübke, "Secure data management in trusted computing," in *Proceedings of the 7th International Workshop on Cryptographic Hardware and Embedded Systems - CHES 2005, Edinburgh, UK, August 29 - September 1, 2005*, ser. Lecture Notes in Computer Science, J. R. Rao and B. Sunar, Eds., vol. 3659. Springer, 2005, pp. 324–338.
- [15] L. Sarmenta, M. van Dijk, C. O'Donnell, J. Rhodes, and S. Devadas, "Virtual monotonic counters and count-limited objects using a TPM without a trusted OS," in *Proceedings of the first ACM workshop on scalable trusted computing*. ACM, 2006, p. 42.
- [16] A. Fuchs, S. Gürgens, and C. Rudolph, "On the security validation of integrated security solutions," in *Emerging Challenges for Security, Privacy and Trust, 24th IFIP TC 11 International Information Security Conference, SEC 2009, Pafos, Cyprus, May 18-20, 2009. Proceedings*, ser. IFIP Advances in Information and Communication Technology, D. Gritzalis and J. López, Eds., vol. 297. Boston, MA, USA: Springer, 2009, pp. 190–201.